ICSA Book Series in Statistics
Series Editors: Jiahua Chen · Ding-Geng (Din) Chen

Karl E. Peace
Ding-Geng Chen
Sandeep Menon *Editors*

Biopharmaceutical Applied Statistics Symposium

Volume 1 Design of Clinical Trials

Springer

Chapter 5
Designing and Analyzing Recurrent Event Data Trials

Stephan Ogenstad

5.1 Introduction

Recurrent event data analysis is common in clinical trials. Literature reviews indicate that most statistical models used for such data are often based on time to the first event or that events within a subject are considered to be independent. Even when taking into account the dependence of the events within subjects, statistical analyses are mostly done with continuous risk interval models, which may not be appropriate for treatments with sustained effects. Furthermore, results can be biased in cases of a confounding factor implying different risk exposure, e.g. in malaria transmission, if subjects are located at zones showing different environmental factors implying different risk exposures (Sagara et al. 2014). Hence, in many prospective randomized controlled clinical trials, events are recurrent, in the sense that the events involve repeat occurrences of the same or different types of events over time. Typical event data consist of times of occurrences of events and the types of events or states that occur. Frequently, an event may be considered as a transition from one state to another and, therefore, multistate models will often provide a relevant framework for event history data. Event history analysis deals with inference for transition intensities and transition probabilities in multistate models. This includes estimation and hypothesis testing for these quantities and analysis of regression models where these quantities are related to explanatory variables observed for the subjects under study. Multistate models are defined by their transition intensities from which transition probabilities may or may not be derived depending on the modeling assumptions. Multistate models are discussed from several points of view in the books and articles by Andersen and Keiding (2002), Andersen et al. (1993), Blossfeld and Rohwer (1995), Courgeau and Lelièvre (1992), and Hougaard (1999, 2000), and Commenges (1999).

S. Ogenstad (✉)
Statogen Consulting LLC, 1600 Woodfield Creek Drive #215, Wake Forest, NC 27587, USA
e-mail: sogenstad@statogen.com

K. E. Peace et al. (eds.), *Biopharmaceutical Applied Statistics Symposium*, ICSA Book Series in Statistics, https://doi.org/10.1007/978-981-10-7829-3_5

The recurrent events are health indicators that assess disease progression or therapeutic effect when subjects are observed over a period of time. It is clinically meaningful to consider whether the treatment a subject is receiving is expected to impact the first event or subsequent events or both. In other words, does the intervention increase the time to the first event or decrease the event number over the study period? In many therapeutic areas, time to the first event is chosen to be the primary endpoint, but this choice then ignores all events after the first one. It is true that statistical approaches for recurrent event endpoints usually are more complex, with less regulatory experience, though there are a number of indications where these endpoints are used, such as hospitalizations in cardiology, asthma and multiple sclerosis. The recurrent event approaches are usually more statistically efficient as information beyond the first event is used. When the follow-up time may be truncated by competing terminal events, it is possible that a subject's observation times may correlate with the competing terminal events themselves, thus making the observation times difficult to assess.

Flexible parametric models of time to the first event or survival can help us in a number of ways. These types of models allow us to obtain estimates of the baseline survival function and its uncertainty which vary smoothly over time. Prediction of survival probabilities and differences, hazard rate functions, hazard differences and ratios, time-dependent effects of covariates, and excess mortality rates in the context of relative survival are just some of the possible outputs from these models.

There are a number of different methods that can be used to evaluate the effects that different treatments can have on subjects in a controlled clinical trial. A few of the methods are to study the 'time to the first event' for the subject, the 'number of events' observed for the subject during the time period that the subject is observed, marginal models, special models with time-dependent covariates, and frailty (random-effects) models. Random effects models are interesting, and our understanding of how they work when applied is beginning to mature. Marginal models are relatively simple to use, interpretable, and flexible, but all of them have limitations. Usually, these models can be fit with standard software such as SAS, Stata and the R package.

In this chapter, we will initially spend some efforts on survival models, since these models form a foundation of many recurrent event models. We are only considering right-censored observations, that is when subjects are still alive at the end of the study and we only have incomplete survival time observations. A crucial problem is whether the available incomplete data enables us to make valid inference on parameters in the multistate model for the complete data. The condition for this is known as independent right-censoring and the interpretation is that a sample observed after independent right-censoring is '*representative*' of the population without censoring. This means that subjects who are censored should have neither lower nor higher risk of future events than subjects who are not censored. We will not cover events that affect trial conduct, such as treatment switching after an event has occurred.

5.2 Methods

5.2.1 Time to First Event

Let T denote a continuous non-negative random variable representing survival time, with probability density function $f(t)$ and cumulative distribution function $F(t) = \mathrm{P}(T \leq t)$. The survival function $S(t) = \mathrm{P}(T > t) = 1 - F(t)$ expresses the probability of a subject being alive at time t. The hazard rate function $\alpha(t) = f(t)/S(t)$ describes the conditional probability of an event occurring at time t, given that the event has not yet occurred. Models based on the hazard rate function can assess whether covariates have an effect on the hazard. If we let $\Lambda(t) = \int_0^t \alpha(u)du$ denote the cumulative or integrated hazard rate function then the survival function can be expressed as $S(t) = \exp(-\Lambda(t))$.

The simplest multistate model is a two-state model where a subject can transition from being '*alive*' (in state 0) to the absorbing state of being '*dead*' (in state 1). Sometimes what is happening to a subject is being viewed as being part of a Markov process. The time it takes until this '*absorbing state*' is reached (or the observational period is censored) is the '*survival time*'. The survival time for a subject will here in the most simple form consist of a random variable, say T, representing the time from a given origin (time 0) to the occurrence of the event '*death*' or we have the knowledge that the observational period is censored. It is seen that $S(t)$ and $F(t)$, respectively, correspond to the probabilities of being in state 0 or 1 at time t. If every subject is assumed to be in state 0 at time 0 then $F(t)$ is also the transition probability from state 0 to state 1 for the time interval from 0 to t. In continuous time the distribution of T may also be characterized by the hazard rate function transition probability from state 0 to state 1 for the time interval from 0 to t. The hazard rate function may be characterized by

$$\alpha(t) = -d \log S(t)/dt = \lim_{dt \to 0} \frac{P(T \leq t + dt | T \geq t)}{dt}$$

that is,

$$S(t) = \exp\left(- \int_0^t \alpha(u)du \right)$$

Thus, $\alpha(.)$ is the transition hazard rate from state 0 to state 1, i.e., the instantaneous probability per time unit of going from state 0 to state 1.

The survival function is often estimated with the Kaplan-Meier (KM) curve (Aalen et al. 2008). It is the most frequently used tool to describe what happened to the subjects in each treatment group. From censored survival data we can easily estimate a survival function by the KM estimator. Figure 5.1 shows KM survival curve estimates

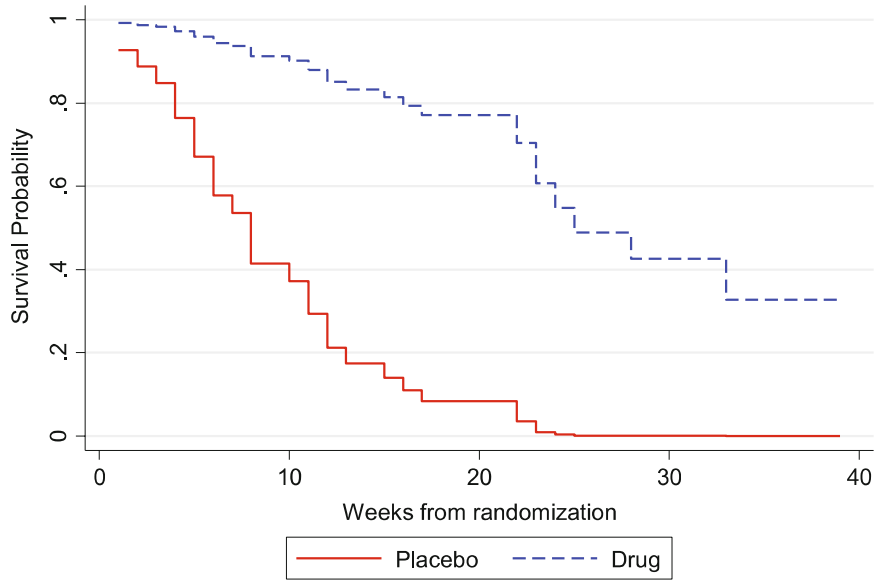

Fig. 5.1 KM survival curve estimates for time to death from cancer for two treatment groups Placebo and Drug

for time to death from cancer for two treatment groups Placebo and Drug. The cancer dataset that ships with the software Stata (cancer.dta) is used, but is entirely fictional.

The least precise parts of the KM curves get the most visual focus, i.e., the right-hand parts of the curves towards the end of the study, where the fewest number of subjects are at risk of the event of death. This is a general criticism of KM survival curve estimates. Kaplan-Meier-type estimates are composed of a sequence of point estimates of the survival functions that are highly serially correlated. Accordingly, KM plots tend to display '*runs*' of values that move away from and back toward the general trend, giving an undulating appearance. This may make the curves difficult to interpret and may lead to the overemphasis of local features (Royston and Lambert 2011).

The estimation of a hazard rate function is more difficult. What can easily be done is to estimate the cumulative hazard rate function $\Lambda(t) = \int_0^t \alpha(u)du$ using the Nelson-Aalen estimator. Figure 5.2 shows the Nelson-Aalen estimates for the same two treatment groups Placebo and Drug described previously.

If the increments of a Nelson-Aalen estimate are smoothed then the new estimates may be used to provide estimates of the hazard rate function themselves. Below are estimates of the hazard rate functions after smoothing of the Nelson-Aalen estimates for the two treatment groups Placebo and Drug (Fig. 5.3).

The smoothing options will, of course, affect the shape of the hazard estimates. We will later on in this chapter show alternative ways of estimating the survival, cumulative hazard and hazard rate functions.

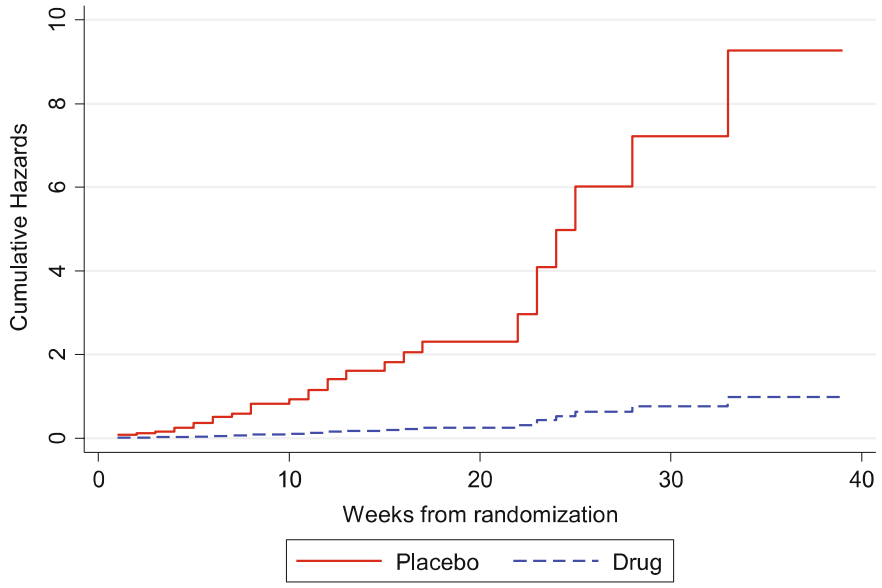

Fig. 5.2 Nelson-Aalen estimates for time to death from cancer for two treatment groups Placebo and Drug

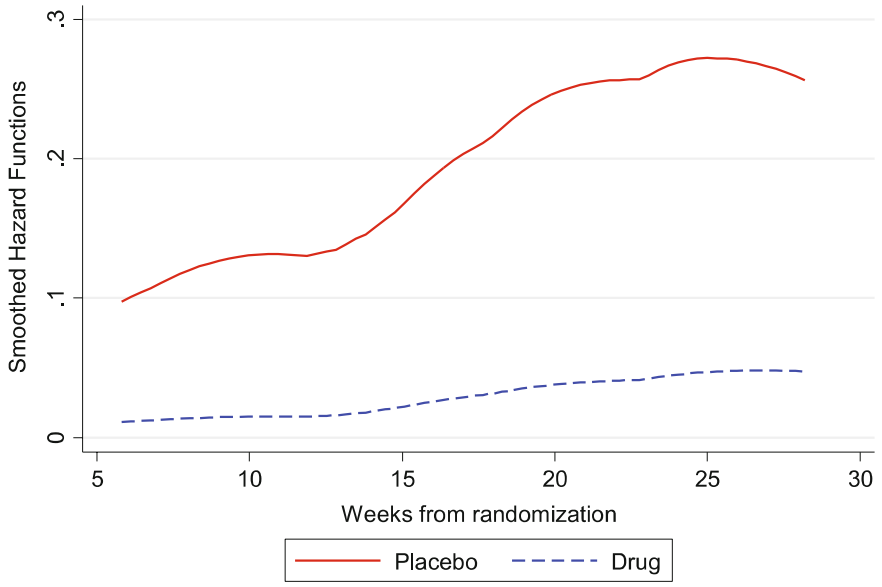

Fig. 5.3 Estimates of the hazard rate functions based on smoothed Nelson-Aalen estimates for time to death from cancer for two treatment groups Placebo and Drug

5.2.1.1 The Cox Proportional Hazards Model

Modeling of censored survival data has since the 1970's almost always been done by the use of the Cox proportional-hazards regression model. The model is in its original form semi-parametric. The hazard rate function for the Cox proportional hazard model (Cox 1972) has the form

$$\alpha(t|\mathbf{z}_i) = \rho_0(t)\exp(\beta_1 z_{i1} + \beta_2 z_{i2} + \ldots + \beta_p z_{ip}) = \rho_0(t)\exp(\mathbf{z}_i'\boldsymbol{\beta})$$

which gives the hazard rate at time t for subject i with covariate vector $\boldsymbol{z}_i$ and parameter vector $\boldsymbol{\beta}$. The baseline hazard $\rho_0(t)$ is arbitrary, which in one sense is scientifically comforting, though the function does not extrapolate any information beyond that. An underlying assumption of the Cox model is that the estimated parameters are not associated with time.

Ignoring ties at the moment and conditioning on the existence of a unique event at some particular time t the probability that the event occurs in subject i for which $C_i = 1$ (uncensored) and $T_i = t$ is

$$L_i(\boldsymbol{\beta}) = \frac{\theta_i}{\sum_{j:T_j \geq T_i} \theta_j}$$

where $\theta_j = \exp(z_j'\boldsymbol{\beta})$. Treating the subjects' events as if they were statistically independent, the joint probability of all realized events conditioned upon the existence of events at those times is the partial likelihood

$$L(\boldsymbol{\beta}) = \prod_{i:C_i=1} \frac{\theta_i}{\sum_{j:T_j \geq T_i} \theta_j}$$

Its log partial likelihood is

$$l(\boldsymbol{\beta}) = \sum_{i:C_i=1} \left(\mathbf{z}_i\boldsymbol{\beta} - \log \sum_{j:T_j \geq T_i} \theta_j \right)$$

This function can be maximized over $\boldsymbol{\beta}$ to produce maximum partial likelihood estimates of the model parameters.

Several approaches have been proposed to handle situations in which there are ties in the time data. The partial likelihood for recurrent failure times is the case when two or more subjects are recorded as dying at the same time. Breslow (1975) developed a method that is the default for many statistical software packages, but it is not the default for the R package. Breslow's method uses the partial likelihood, expressed as

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{I} \frac{\prod_{j \in D(t_{(i)})} \phi_j}{\left(\sum_{j \in R(t_{(i)})} \phi_j\right)^{|D(t_{(i)})|}}$$

where $\left|D(t_{(i)})\right|$ is the number of subjects that fail at time $t_{(i)}$.

Breslow's method describes the approach in which the procedure described above is used unmodified, even when ties are present. An alternative approach that is considered to give better results is Efron's method (Efron 1974). The Cox model may be specialized if a reason exists to assume that the baseline hazard follows a particular form. In this case, the baseline hazard $\rho_0(t)$ is replaced by that particular function. An alternative to Cox's model is the additive regression model due to Aalen (Aalen et al. 2008), which assumes that the hazard rate of a subject i with p covariates z_{i1}, …, z_{ip} takes the form

$$\alpha(t|\mathbf{z}_i) = \beta_0(t) + \beta_1(t)z_{i1} + \ldots + \beta_p(t)z_{ip}.$$

For this model $\beta_0(t)$ is the baseline hazard, while the *regression functions* $\beta_j(t)$ describe how the covariates affect the hazard rate at time t. For the Cox and additive regression model hazard rate functions, the covariates are assumed to be fixed over time. More generally, one may consider covariates that vary over time (Aalen et al. 2008). The generic term parametric proportional hazards models can be used to describe proportional hazards models in which the hazard rate function is specified. The Cox proportional hazards model is sometimes called a semiparametric model by contrast.

The R package uses Efron's partial likelihood, as it is considered a closer approximation to the exact partial likelihood. Efron's partial likelihood has the following shape

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{I} \frac{\prod_{j \in D(t_{(i)})} \phi_j}{\prod_{k=1}^{|D(t_{(i)})|} \left(\sum_{j \in R(t_{(i)})} \phi_j - \frac{k-1}{|D(t_{(i)})|} \sum_{j \in D(t_{(i)})} \phi_j\right)}$$

An extension of the proportional hazards model is to allow for multiple strata in the fitting procedure. That is, we assume that the subjects can be broken into multiple groups, and the hazard rate function for subjects in the kth group is

$$\rho_{0k}(t) \exp(\mathbf{z}_i' \boldsymbol{\beta}).$$

A common use of stratification is in multicenter trials. Because of different subject populations and referral patterns, different centers in the trial may have quite different hazard rates, yet a common treatment effect across centers. In this way, strata play a similar role to multiple intercept terms in an analysis of covariance model. Each baseline hazard captures the baseline rate for an event. When events are of different types, we have in reality different baselines. If we, for instance, are studying heart

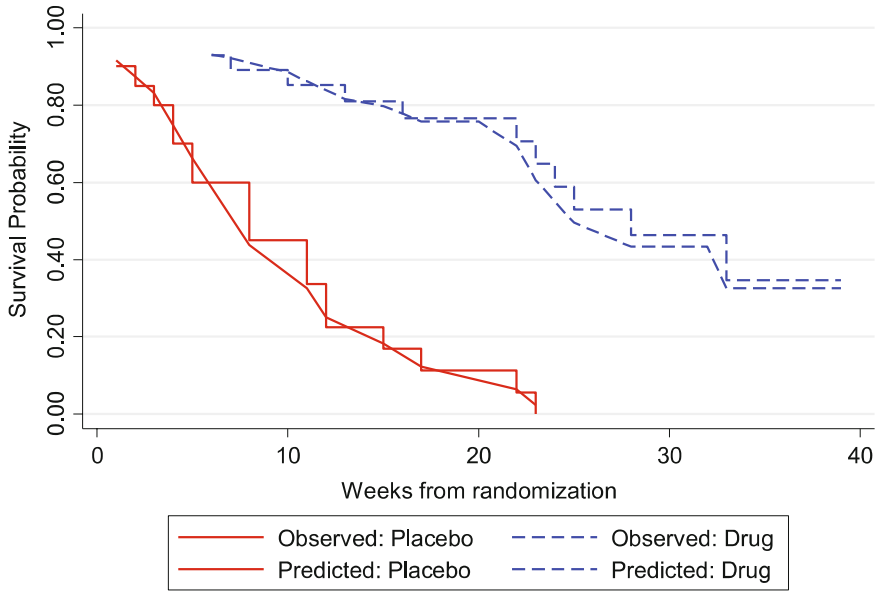

Fig. 5.4 KM and Cox model estimates for time to death from cancer for two treatment groups Placebo and Drug

attacks, are the first and second heart attacks the same type of event? Well, we would only know if we investigate it.

The Kaplan-Meier and Cox estimation provide estimates of the survival functions. Continuing to use our cancer data, Fig. 5.4 displays the KM and Cox model estimates.

5.2.1.2 Extending the Cox Model for the Two-State Case

The main purpose of the Cox model in its simplest form is to estimate hazard rates assuming that the hazards are proportional to each other. Because the model can be embedded in a counting process framework (Andersen et al. 1997), the model can be extended in many different ways to answer questions across a wide range of situations, where we need to obtain informative estimates of quantities that include hazard rates and their differences and ratios, survival curves and their differences, rates, and survival at given time points. By '*informative*' we mean unbiased estimates that are smooth functions.

Parametric survival models generally provide smooth estimates of the hazard and survival functions for any combination of covariate values. The exponential model is often used when planning a clinical trial and for calculating the power and sample sizes. Though, the exponential survival model is a rather unrealistic model since it is assumed that the hazard rate function is constant over the whole observational study period. This model can be generalized by splitting the observational period into intervals. The choice of the number of intervals and where to place the cutpoints

is of course subjective. With the piecewise exponential model, the time scale is split into several intervals, where we assume that the hazard rate function is constant within each interval but can vary from interval to interval. The hazard rate function for the piecewise exponential model can be written $h_{ij}(t, \mathbf{z}_i) = \alpha_j \exp(\mathbf{z}_i'\boldsymbol{\beta})$, where the subscript i is for subject and j is for the interval. Modeling the data with the Poisson approach allows us to think about survival time in a different way from that in standard survival analysis. Usually, survival time is considered to be the outcome variable and we have to use special methods to account for the censoring process. With the Poisson approach, it becomes clearer that we are modeling rates. We have a binary variable as an outcome, and our models investigate variation in the corresponding rates. There are many factors that cause systematic variation in rates, for example, age and gender, but also time. In the Poisson framework, we can, therefore, consider time to be a covariate, as opposed to a response. Thus we can adjust for time just as we would for any other covariate. Time-dependent effects of a covariate of interest are then simply an interaction between time and the covariate. The Poisson model with a split at each unique failure time gives us the Cox model. However, we do not want to fit a model with so many parameters. An important question is, what is the effect of changing the number of time intervals of the parameters of interest (usually log hazard ratios)? The problem with the piecewise exponential model is that if we choose too few intervals we may miss important changes in the hazard rate; if we choose too many, we end up with too many parameters, and the underlying shape of the hazard rate is difficult to see because of random variation.

5.2.1.3 Royston-Parmar Models

The use of parametric models for the type of data we so far have considered may have some advantages. The non-proportional hazards that are a potential difficulty with the Cox model, could sometimes be handled in a simpler way, and the visualization of the hazard rate function could be much easier. Royston–Parmar (RP) models (Royston and Parmar 2002, Lambert and Royston 2009) have great flexibility with respect to the shapes of the survival distributions they can model. Familiar standard parametric survival models are the starting point for the generalizations called RP models. Weibull, log logistic, and lognormal models can be generalized to proportional hazards, proportional odds, and probit-scaled RP models, respectively. The additional flexibility of RP models arises because the baseline distribution function is represented as a restricted cubic spline function of log time instead of simply as a linear function of log time. Modeling with spline functions generates some additional complexity. The additional complexity is determined by the number and the positions of the connection points in log time, known as *knots*, of the spline's cubic polynomial segments. Estimation of parameters is by maximum likelihood. Quite often, the characteristics of the fitted model are rather insensitive to the number and particularly the position of the knots, lending a certain robustness to the process of model selection. The restriction that the transformed survival function be linear in Ln(t) is, in practice, severely limiting and is not really necessary. In RP models, we may relax linearity and allow nonlinear functions. There are many possible fami-

lies of nonlinear functions that we could use. Because cubic splines are flexible yet relatively simple to work with and understand, Royston and Parmar (2002) chose them as their preferred tool to extend standard models. The result is a major advancement in the practical usefulness of parametric survival analysis and in the range of applications that can be tackled.

In cancer survival trials, one often wants to know the impact of covariates on the mortality rate for a particular cancer diagnosis. Since cancer is mostly a disease of old age, many people may die of diseases other than the specific type of cancer they were originally diagnosed with. Relative survival is a measure of patient survival corrected for the effect of other causes of death by utilizing the patients' expected survival. Both Poisson models and Royston–Parmar (RP) models can be extended to relative survival by incorporating information on expected survival or mortality. Relative survival is related to the concept of competing risks (Gamel and Vogel 2001). We there assume that an individual is at risk of either dying of their cancer or dying of another cause. In relative survival models, we can deal with this issue by incorporating expected mortality, which can usually be obtained from routine data sources. Traditionally, simple piecewise models have been used for relative survival, but all the advantages of standard parametric survival models also apply to relative survival models.

The baseline survival function in a Cox model is available only in the estimation sample. To predict survival outside the estimation sample, we need special measures, such as interpolation or even extrapolation. Using special measures limits the applications of the Cox model in some situations. An important case arises when we wish to validate a survival model in an independent sample, a task that necessitates out-of-sample prediction. There are at least two situations in which this is useful. One is by interpolating or extrapolating the baseline or other survival functions at time points not represented in the estimation sample. The other is by predicting survival probabilities or other quantities of interest from a model on a derivation sample onto individuals in an evaluation sample (that is, external validation). Interpolation is helpful, for instance, when we wish to plot a survival function for an individual, a group, or a covariate pattern as a smooth curve at a suitable choice of time points within the range of the observed follow-up time. We need the extrapolation when we want to project a modeled survival function into the future. Successful external validation is usually regarded as the gold standard of potential usefulness of a proposed prognostic model (Altman and Royston 2000).

The Hazard rate function is of utmost relevance in clinical medicine since it is a decidedly meaningful measure of disease course, and is the basis against which relative hazard effects are estimated. Fuchs et al. (1994) report on a double-blind randomized multicenter clinical trial designed to assess the effect of rhDNase (purified recombinant form of the human enzyme DNase I) versus placebo on the occurrence of respiratory exacerbations among patients with cystic fibrosis. The subjects in these treatment groups are susceptible to an accumulation of mucus in the lungs, which leads to pulmonary exacerbations and deterioration of lung function. The occurrences of exacerbations over the study period were recorded for each subject. The estimated hazard functions in the Fig. 5.5 are derived from the Cox model and the

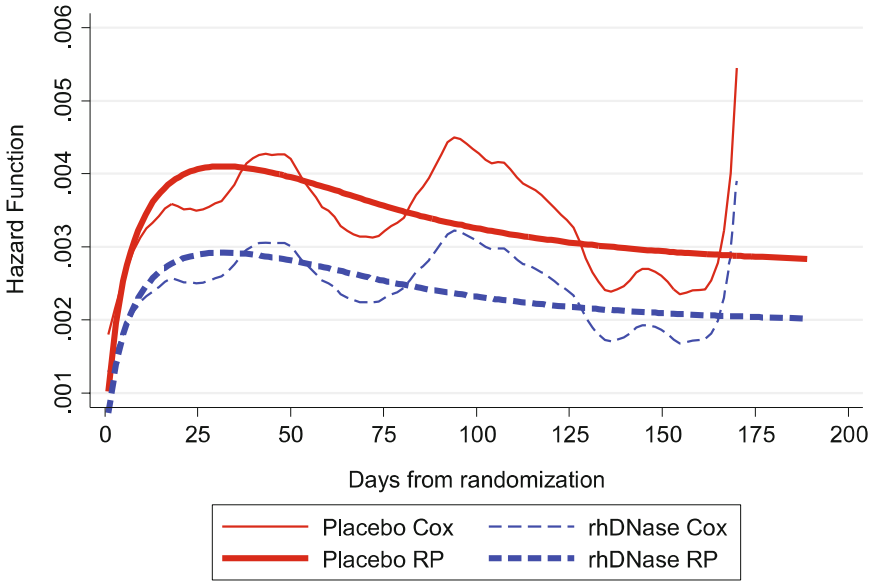

Fig. 5.5 Cox model and the Royston-Parmar model estimates under the proportional hazards assumption from rhDNase data

Royston-Parmar model under the proportional hazards assumption. The thicker pair of lines show estimates of the hazard rate functions from the Royston-Parmar model and the thinner lines from the Cox model, using kernel smoothing, in the Placebo and rhDNase groups, respectively.

The Royston-Parmar model gives a more plausible trajectory of the hazard for the patients, then the rugged course that is shown from the estimates based on the Cox model. In the RP model, the hazards seem to be highest about one month after randomization and decreases after that time. The hazards are substantially reduced by the rhDNase treatment. The proportional hazards condition forces the curves to be proportional to each other. Even after 175 days the hazard in the rhDNase treatment arm is still substantial but reduced by about one third. The fact that the curve does not approach zero suggests that the disease is chronic. We have obtained quite a lot of useful information. Even if we relax the proportional hazards assumption, the plot of the ensuing hazard rate functions (not shown) are very similar to the thick lines in the figure. So, our conclusion about the treatment effect seems to be robust.

The baseline hazard contains useful information. If we are told that the mortality rate is double for subjects with a particular exposure, then we want to know what reference value this doubling refers to. In a survival model, the reference is usually the baseline hazard rate, which usually changes as a function of time. Thus even if the proportional hazards assumption is reasonable, the impact of a particular exposure in absolute terms depends on how long time has passed since the time origin (diagnosis, randomization, start of treatment, etc.) and the magnitude of the underlying hazard

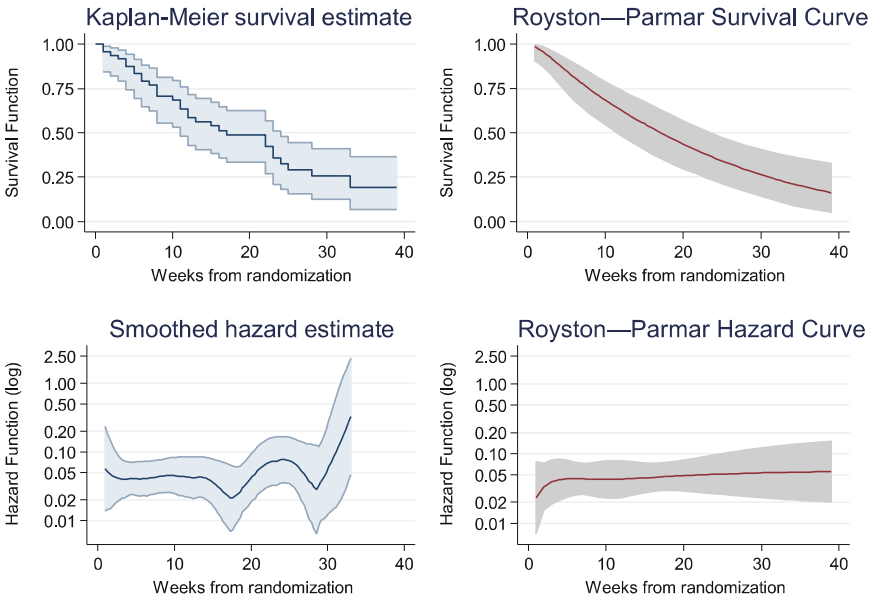

Fig. 5.6 Estimated survival and hazard rate functions with confidence intervals, for time to death from cancer for two treatment groups Placebo and Drug

rate. Flexible parametric survival models can help us in a number of ways. For instance, these models allow us to obtain an estimate of the baseline survival function and its uncertainty which vary smoothly over time.

We will illustrate (Fig. 5.6) the survival distributions and hazard functions using nonparametric techniques (Kaplan–Meier and smoothed hazard functions, respectively) and a flexible parametric technique (Royston–Parmar models) using the cancer dataset.

The survival curves indicate a median time to event of about 16–20 weeks. The Kaplan–Meier curve shows a slight downturn after about 22 weeks, which is not reflected in the survival curve from the Royston-Parmar estimates. The smoothed nonparametric hazard estimate shows a corresponding upturn about 30 weeks. Whether the feature is "real" or not is questionable—it seems surprising that the event rate would start to increase after 18 weeks and then gradually turning down at 30 weeks, and then again from thereon shoot up. The pointwise confidence intervals (CIs) from the smoothed hazard estimate are wider than that from the Royston-Parmar. Conditional on a parsimonious parametric model, CIs are generally too narrow because they do not take model uncertainty into account. Nonparametric CIs make fewer assumptions and tend to be wider. Also, they are implicitly high-dimensional and noisy.

The Royston-Parmar models can equally be used to perform multistate survival analysis (Crowther 2016).

5.2.2 *Multiple Events Per Subject*

5.2.2.1 Multistate Models

We have considered situations where each subject can only have one event. If death is the outcome, then clearly it is not possible to have more than one event. However, if the event is the recurrence of disease or readmission to hospital, then it is possible for each subject to have more than one event. As a continuation of survival analysis, we will consider another type of multivariate data in the setting of competing risks, where $T_1, T_2, \ldots, T_k$ represent survival times to different causes of death. Estimation of these models is complicated by the fact that we only observe $T = \min\{T_1, \ldots, T_k\}$ where even T can be censored, so that none of the k events have occurred. Yet, another type of multivariate data involves transitions among several types of states, where some of them might be terminal, but not all. This combines elements of competing risk models with models for series of events.

The framework for these types of models can be set up in the following way: Suppose there is a total of m subjects accrued into a study and each subject is at risk for a particular type of recurrent event. Let $(0,\tau]$ represent the period of observation and let $N_i^*(u)$ be a right-continuous integer function representing the number of events experienced by subject i over the interval $(0, u]$, $i = 1, 2,\ldots, m$, $0 < u \leq \tau$. During the observation period $(0, \tau]$, some subjects may experience an event which terminates their recurrent event processes (e.g. death), but subjects may also withdraw from the study according to some random censoring mechanism which is independent of the recurrent event and terminal event processes. For $i = 1, \ldots, m$, let T_i be the time of the terminating event, C_i the censoring time, $X_i = \min(T_i, C_i)$, $\pi(t) = \mathrm{P}(X_i \geq t)$, and $\delta_i = I(X_i = \mathrm{d}T_i)$, where $I(\cdot)$ is an indicator function. We let $N_i(t) = N_i^*\{\min(t, T_i)\}$ denote the number of recurrent events observed over $(0, t]$ in the presence of death. The data contributed by each subject then take the form $(\{N_i(u), 0 < u \leq X_i\}, X_i, \delta_i)$, $i = 1,\ldots, m$. Let $Y_i(t) = I(X_i > t)$ be the at risk indicator function which is one when subject i is under observation and at risk for an event at time t and is zero otherwise. We suppose initially that we have a single sample of subjects.

The most important class of models is the continuous time Markov process $X(t)$ on the finite state space $S = \{1,\ldots, p\}$ where the dependence of transition hazard rate function $\alpha_{nj}^i(t)$ on the history X_t is only through the current state of $X(t)$ and possibly via time-fixed covariates. Statistical models are usually obtained by specifying the class of transition intensities $(\alpha_{nj}^i(t))$ for each subject i.

The most important deviations from the Markov property in practice are various kinds of duration dependence, where transition intensities depend on other time origins than $t = 0$, typically the time of entry to the present state. There are two main approaches to handling these. As long as transition intensities depend only on one-time origin each (for example, all intensities depend only on duration in the present state), a model for the multistate process may be obtained by combining independent submodels for each transition hazard rate. These may, in turn, be modeled as constant or piecewise constant or by non- or semiparametric models, and as long as

there is a unidirectional flow in the model, transition probabilities are still straightforward explicit functionals, which may be estimated by plugging in the hazard rate estimates. Variance calculations may, however, become less direct (Andersen and Keiding 2002).

5.2.2.2 Univariate Recurrent Events

At the moment, we are only concerned with univariate events, i.e., events of the same kind. Because the various events occur to the same subject, the waiting times will in general not be independent. Since the events occur one after the other, it will generally be the case that only the last interval can be censored.

With recurrent events, we can expect a correlation between the times to event of a given subject. For instance, subjects with severe disease will tend to have more events and a shorter time between events than those with mild disease. The most common models used are (i) Generalized estimating equations model using a Poisson or Negative Binomial distribution, and three extended Cox models: (ii) the Andersen-Gill counting process (AG) (Andersen and Gill 1982), (iii) the Prentice-Williams-Peterson counting process (PWP) (Prentice et al. 1981), (iv) Wei, Lin, and Weissfeld (WLW) (Wei et al. 1989; Lin 1994) and (v) the frailty model (Gutierrez 2002). For the marginal models, the correlation is dealt with using a robust sandwich-based estimator to avoid inflation of type I error due to multiple observations per subject which do not require specification of the correlation matrix (Kelly and Lim 2000). Consideration needs to be taken whether the events are ordered or not. Ordered events could be, for instance, first, second, third, … hospitalization. Unordered events could, for instance, be of different types, such as '*hospitalization*', '*withdrawal*', and '*death*', where '*death*' is a competing event. For more details of the approaches, see Therneau (1997) and Therneau and Grambsch (2000). One can fit similar models within the Royston-Parmar framework (Royston and Parmar 2002; Lambert and Royston 2009).

5.2.3 Poisson Regression

A Poisson process can be described via the hazard rate function that is of the form

$$\alpha(t|H(t)) = \rho(t) \quad t > 0,$$

where $\rho(t)$ is a nonnegative integrable function. It is also assumed that the cumulative hazard rate

$$\mu(t) = \int_0^t \rho(u)du \quad t > 0,$$

is continuous and finite for all $t > 0$. It is seen from the hazard rate function above that the Poisson process is Markovian. The probability of an event in $(t, t + \Delta t)$ may depend on t but is independent of the history $H(t)$.

Poisson regression is a generalized linear form of regression analysis used to model count response data. Poisson regression assumes the response variable Y has a Poisson distribution and assumes that the logarithm of its expected value can be modeled by a linear combination of unknown parameters. The Poisson regression model is frequently used to analyze count data when the dependent variable represents the number of independent events that occur during a fixed period of time (Prentice et al. 1981, Sagara et al. 2014). The method assumes that all events are independent and is based on event rates, where the total number of events is divided by the follow-up time. The conditional mean of Y (the number of events) can be written as:

$$Ln(Y|\mathbf{Z}, \boldsymbol{\beta}) = \mathbf{Z}_i\boldsymbol{\beta}$$

where $\mathbf{Z}_i\boldsymbol{\beta} = \beta_0 + \beta_1 Z_1 + \ldots + \beta_k Z_k$ of k parameters and Ln is the natural logarithm function.

The probability function for a unit-time interval for a subject i can be expressed as

$$f_Y(y_i; \mu_i) = e^{-\mu_i}\mu_i^{y_i}/y_i!$$

for $y = (0, 1, \ldots)$ and $\mu_i > 0$. The mean and variance are both equal to μ_i. With subscripts indicating subject i's observation the log-likelihood function can be written as

$$L(\mu_i; y_i) = \sum [y_i \ln(\mu_i) - \mu_i - \ln(y_i!)]$$

The parameter μ_i can be reparameterized as $\exp(\mathbf{z}_i'\boldsymbol{\beta})$, and therefore the log-likelihood function can be written as

$$L(\boldsymbol{\beta}_i; y_i) = \sum [y_i(\mathbf{z}_i\boldsymbol{\beta}) - \exp(\mathbf{z}_i\boldsymbol{\beta}) - \ln(y_i!)]$$

5.2.4 *Negative Binomial Regression*

One of the key features of the Poisson distribution is that the variance equals the mean. However, one often finds that overdispersion is frequent in count data. Overdispersion in a Poisson model occurs when the variance of the response is greater than the mean. One approach to handling the overdispersion is to add covariates to the model. Though, even after conditioning on covariates, there could still be more inter-subject variation in event occurrence than accounted for by a Poisson process. Another approach is then to model the overdispersion by adding a multiplicative random

effect to represent unobserved heterogeneity. Doing so will lead to the negative binomial regression model where the conditional distribution of the outcome Y, given an unobserved variable θ, is indeed Poisson with mean and variance $\theta\mu$. The variable θ captures unobserved factors that increase (if $\theta > 1$) or decrease (if $\theta < 1$) relative to what we would expect given the observed values of the covariates. In this model, the data would be Poisson if only we could observe θ. Unfortunately, we do not. Instead, we make an assumption regarding its distribution and integrate θ out of the likelihood, effectively computing the unconditional distribution of the outcome. It is mathematically convenient to assume that θ follows a gamma distribution. The unconditional distribution of the outcome is the negative binomial distribution (Cook and Lawless 2007; Hilbe 2007).

5.2.5 *Extended Cox Models for Recurrent Events*

As we mentioned, recurrent event data are correlated since multiple events may occur within the same subject. While using frailty models is one method to account for the correlation in recurrent event analyses, a simpler approach that can also account for this correlation is the use of robust standard errors (SEs). With the addition of robust SEs, recurrent event analysis can be done as a simple extension of either semi-parametric or parametric models.

If interest focuses on recurrent occurrences of a given event, for instance, hospitalization, then another model than the Cox model should be considered. In applications of such a model, an interesting functional is often the expected number of occurrences of the event over the time interval $(0, t]$. The corresponding semi-parametric estimate of the cumulative expected number of events over $(0, t]$ for subject i is

$$\hat{E}[N_i(t)] = \int_0^t \hat{\rho}_0(s) \exp(\mathbf{z}_i' \hat{\boldsymbol{\beta}}) ds$$

where N_i(t) is the number of events for subject i over $(0, t]$. This is the same as the generalized Nelson–Aalen, or Breslow, estimate from survival analysis. (Cook and Lawless 2002; Andersen et al. 1993).

Cumulative Sample Mean Function

Plots like the one in Fig. 5.7 have limitations since it is often not easy to determine visually whether a trend or other patterns exist in data.

A visually more informative function is the cumulative sample mean function (Cook and Lawless 2007). The function can be defined as follows. Suppose that m individual processes are observed, with each process being observed over the time interval $(0, t]$. Let $N_i(t)$ represent the number of events over the time interval $(0, t]$ for the ith process. Then the cumulative sample mean function is

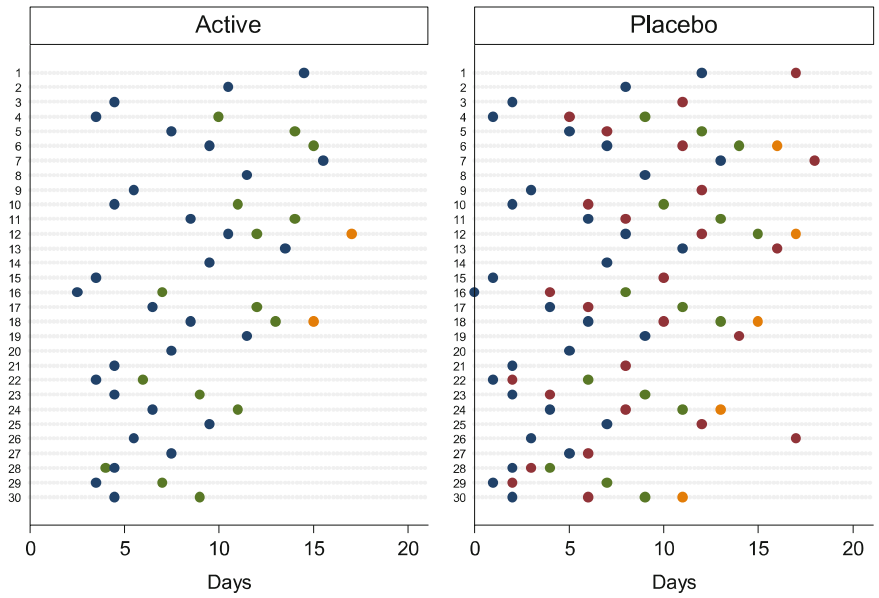

Fig. 5.7 Event plots from time of randomization for tumor occurrence in 60 subjects (30 subjects on Active and Placebo, respectively)

$$\hat{\mu}(t) = \frac{1}{m} \sum_{i=1}^{m} N_i(t).$$

The same data as in Fig. 5.7 is used to display the cumulative sample mean function (Fig. 5.8).

5.2.5.1 The Andersen-Gill Model (AG)

The counting process, or Andersen-Gill, approach to recurrent event modeling assumes that each recurrence is an independent event, and does not take the order or type of event into account. In this model, follow-up time for each subject starts at the beginning of the study and is broken into segments defined by events (recurrences). Subjects contribute to the risk set for an event as long as they are under observation at that time (not censored). The model is simple to fit as a Cox model with the addition of a robust standard error estimator, and hazard ratios are interpreted as the effect of the covariate on the recurrence rate over the follow-up period. This model would be inappropriate, however, if the independence assumption is not reasonable.

External covariates $x(t)$, which include fixed covariates, can be incorporated in a Poisson process by specifying the hazard rate as a function of t and the covariate history $x^{(t)} = \{x(u) : 0 \leq u \leq t\}$. This is usually done by defining covariate vectors

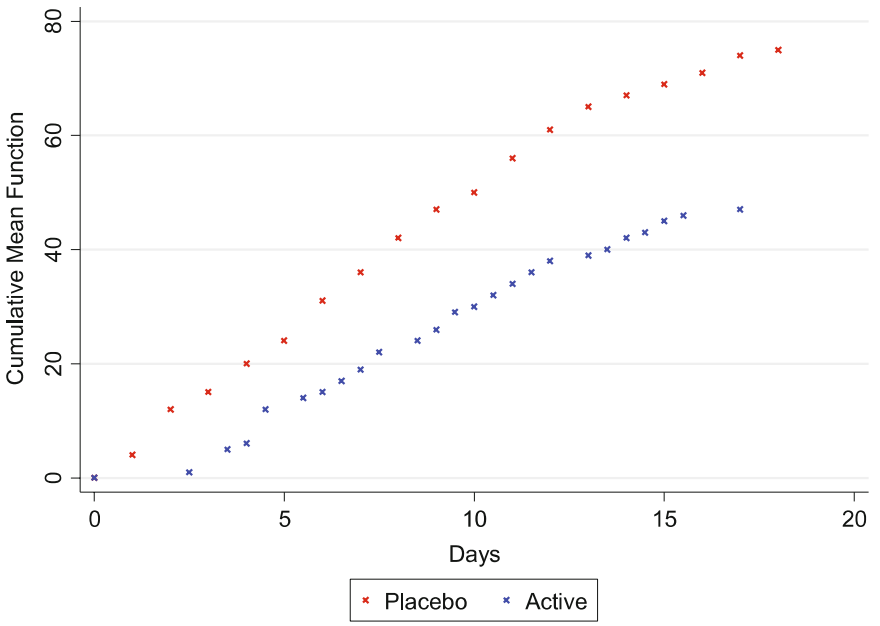

Fig. 5.8 The cumulative sample mean function from time of randomization for tumor occurrence in 60 subjects (30 subjects on Active and Placebo, respectively)

$\mathbf{z}(t)$ that are based on $x^{(t)}$ and then considering the multiplicative intensities of the form

$$\rho(t|x^{(\infty)}) = \rho(t|x^{(t)}) = \rho_0(t)\exp(\mathbf{z}'(t)\boldsymbol{\beta}),$$

where $\boldsymbol{\beta}$ is a vector of regression parameters of the same length as $\mathbf{z}(t)$. The positive valued function $\rho_0(t)$ is often called the baseline rate or intensity and corresponds to a subject for whom $\mathbf{z}(t)=0$ for all $t>0$. This model is sometimes referred to as a log-linear model. The exponential term can be replaced by a different positive term but has been chosen for mathematical convenience. When the baseline function $\rho_0(t)$ is specified as nonparametric then the model is semiparametric and is called the Andersen-Gill (AG) (1982) model.

The AG model is an extension of the Cox model and uses the counting process timescale for all events. The time-scale does not reset to 0 after an event but continues from the time point of the event. Data for each subject needs to be entered in the counting process style, with a start time, stop time and censoring indicator for each event. The model is close in spirit to Poisson regression and the increments are assumed to be independent. Each gap time (interval from one event to the next) contributes to the likelihood and the model assumes that the events are independent. The AG model splits the time scale where the split points are defined by the time point when the events occur. The time intervals are non-overlapping; that is, the start

time of a new event is the ending time of the preceding event. In the AG model, the underlying shape of the baseline hazard is assumed to be the same for all events; that is, there is no stratification by event number. Although not specified in the original article, cluster-based robust standard errors are usually used.

The sandwich robust standard error of Lin and Wei (1989) which is a variance-correction technique, is usually employed together with these Cox extended models to avoid inflation of type I error due to multiple observations per subject which do not require specification of the correlation matrix.

5.2.5.2 Conditional Counting Process Model by Prentice-Williams-Peterson (PWP)

The PWP model is a conditional model, similar to the AG model, but stratified by events. The hazard rate function is written as:

$$\rho_{ik}(t|x^{(t)}) = \rho_{0k}(t)\exp(\mathbf{z}'_{ik}(t)\boldsymbol{\beta})$$

$\rho_{0k}(t)$ represents the event-specific baseline hazard for the kth event over time. In this model, a subject is assumed not to be at risk for a subsequent event until a current event has terminated. The PWP model is similar to the AG model in that it uses nonoverlapping time intervals (gap times) for each subject. As for the AG model, it is not possible to be at risk of the second event before the first event has occurred. The PWP model differs from the AG model in that the baseline hazard for each event k is allowed to be different; that is, there is stratification by event number.

5.2.5.3 The Wei, Lin, and Weissfeld (WLW) Model

Suppose there are n subjects and each subject can experience up to K potential events. Let $\mathbf{Z}_{ki}(t)$ be the covariate process associated with the kth event for the ith subject. The marginal Cox model is given by

$$\rho_{ik}(t|x^{(t)}) = \rho_{0k}(t)\exp(\mathbf{z}'_{ik}(t)'\boldsymbol{\beta}_k),\ k = 1,\ldots,K; i = 1,\ldots,n$$

$\rho_{0k}(t)$ is the (event-specific) baseline hazard function for the kth event and $\boldsymbol{\beta}_k$ is the (event-specific) column vector of regression coefficients for the kth event. The WLW model estimates $\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_K$, by the maximum partial likelihood estimates $\hat{\boldsymbol{\beta}}_1,\ldots,\hat{\boldsymbol{\beta}}_K$, respectively, and uses a robust sandwich covariance matrix estimate for $(\hat{\boldsymbol{\beta}}'_1,\ldots,\hat{\boldsymbol{\beta}}'_K)'$ to account for the dependence of the multiple failure times.

The WLW model uses overlapping time intervals for each subject and stratum so that each stratum is fit separately, and then the estimates are combined. This implicitly forces all strata and covariate interactions to be present. This is equivalent to fitting all of the data at once, i.e., the events are occurring in parallel. The model treats an

ordered dataset as though it were an unordered dataset in a competing risks problem. Thus, each event or event type is in its own stratum and all time intervals starting at 0. Hence, the WLW approach considers each event to be a separate process, so subjects are at risk for all events from the start of follow-up, regardless of whether they experienced a prior event. This model is appropriate when the events are thought to result from different underlying processes, so that a subject could experience the 3rd event, for example, without experiencing the 1st. Although this assumption seems implausible with some types of data, like cancer recurrences, it could be used to model injury recurrences over a period of time, when subjects could experience different types of injuries over the time period that have no natural order. There is a need to specify the total number of events in advance. The method also analyzes the gap times between different events.

5.2.5.4 Competing Risks

Traditional survival analysis methods assume that only one type of events of interest occurs. A way to avoid dealing with competing events in a more complex model than the Cox model is to construct composite endpoints. An example of this is when studying cardiovascular outcomes in type 2 diabetes, where the primary composite outcome is the time-to-event of the first occurrence of death from cardiovascular causes, nonfatal (including silent) myocardial infarction, or nonfatal stroke (Marso et al. 2016). Models in which there are different types of events (*multiple destinations*) are also of interest. Competing risks occur when a subject is at risk of more than one type of event, but can actually experience only one of them. The most common case is when the different events are death from different diseases, such as cancer, heart disease, or an infection. Competing risk models are a special case of multistate models in which each of the different events are absorbing states (Andersen et al. 2002). In competing risks, a subject is at risk of dying from one of, say K, different causes, but can only actually die of one cause.

More complex methods exist to allow the investigation of several types of events in the same study, such as death from multiple causes. Competing risks analysis is used for these studies in which the survival duration is ended by the first of several events. Special methods are needed because analyzing the time to each event separately can be biased. Specifically, in this context, the Kaplan-Meier method tends to overestimate the proportion of subjects experiencing events. Competing risk analysis utilizes the cumulative incidence method, in which the overall event probability at any time is the sum of the event-specific probabilities. The models are generally implemented by entering each study participant several times—one per event type. For each study participant, the time to any event is censored on the time at which the patient experienced the first event.

The two most significant measures in competing risks are the cause-specific hazard rate and the cumulative incidence function. The cause-specific hazard rate function for cause k, say $h_k(t)$, gives the hazard rate at time t conditional on not having died of any of the K possible causes of death. The cause-specific hazard, $h_k(t)$, can be

estimated by treating events due to competing causes as censored observations. The K cause-specific hazard rates are usually estimated by fitting K separate models or by stacking the events (having K rows of data per subject) and fitting a model stratified by cause (Lunn and McNeil 1995). The second most important measure is the cumulative incidence function, say $C_k(t)$, for the kth competing event. This gives the probability, as a function of time, that a subject dies of cause k in the presence of competing risks. It recognizes that a subject cannot die of cause k if that subject has already died of one of the competing causes. The cumulative incidence function is also known as the crude probability of death (Tsiatis 2005). It can be contrasted with the net probability of death, which gives the probability of dying in a situation where it is impossible to die of other causes. cumulative incidence functions give probabilities of death where subjects are always at risk of death from several different causes. The cumulative incidence is calculated from a relative survival model and is defined as

$$C_k(t) = \int_0^t h_k(u) \exp\left\{ - \int_0^u \sum_{k=1}^{k} h_k(v) dv \right\} du = \int_0^t h_k(u) \prod_{k=1}^{K} S_k(u) du.$$

$C_k(t)$ can be calculated by using the Stata package (Fine and Gray 1999).

A nonparametric analysis of recurrent events in the presence of death as a competing risk has been developed by Ghosh and Lin (2000) and by Li and Lagakos (1997).

5.2.5.5 Period Analysis

Cancer survival measures the effectiveness of health-care systems. Persistent regional and international differences in survival represent a source of information that may be used to avoid early death. Differences in survival have impelled or steered cancer control strategies. Statistics reflective of patient survival should be as current as possible. The traditional methods for analyzing survival have important shortcomings with regard to how current they are with respect to long-term cumulative survival estimates. An alternative approach denoted '*period analysis*', that may be used to overcome or reduce these constraints. When cancer survival is improving over time, the use of older data underestimates the survival proportion. One potential solution to this is to use period analysis to obtain more up-to-date estimates of patients' survival (Brenner and Gefeller 1997). This approach has become widely established in the analysis of population-based cancer survival. For example, it has been used in a number of recent international comparisons of cancer survival (Coleman et al. 2011; Møller et al. 2010). Period estimates of patient survival are usually calculated separately in subgroups of interest using life table methodology. Up-to-date estimates of patient survival using period analysis are based on artificially truncating individuals' survival times prior to a recent cutoff in calendar time. This has the effect of using

individuals diagnosed in a recent time period for short-term survival and individuals diagnosed further back in time for longer term survival.

In the Coleman et al. (2011) study, data from population-based cancer registries in 12 jurisdictions in six countries were provided for 2.4 million adults diagnosed with primary colorectal, lung, breast (women), or ovarian cancer during 1995–2007, with follow-up to Dec 31, 2007. Data quality control and analyses were done centrally with a common protocol, overseen by external experts. They estimated 1-year and 5-year relative survival, constructing 252 complete life tables to control for background mortality by age, sex, and calendar year. Also, they reported age-specific and age-standardized relative survival at 1 and 5 years, and 5-year survival conditional on survival to the first anniversary of diagnosis. In addition, they examined incidence and mortality trends during 1985–2005. Their findings were that relative survival improved during 1995–2007 for all four cancers in all jurisdictions.

In the Møller et al. (2010) study, several international studies reported that survival from breast cancer is lower in the United Kingdom than in some other European countries. They compared breast cancer survival between the national populations of England, Norway, and Sweden, with a view to identifying subsets of patients with particularly good or adverse survival outcomes. They also extracted cases of breast cancer in women diagnosed 1996–2004 from the national cancer registries of the 3 countries. The study comprised 303,657 English cases, 24,919 Norwegian cases and 57,512 cases from Sweden. Follow-up was in 2001–2004. The main outcome measures were 5-year cumulative relative survival and excess death rates, stratified by age and period of follow-up.

5.2.5.6 Frailty Models

Correlated survival data can arise due to recurrent events experienced by an individual or when observations are clustered into groups. Either due to lack of information or for feasibility, some covariates related to the event of interest may not be measured. Frailty models account for the heterogeneity caused by unmeasured covariates by adding random effects that act multiplicatively on the hazard function. Frailty models are essentially extensions of the Cox model with the addition of random effects. Although there are various classification schemes and designation used to describe these models, four common types of frailty models include shared, nested, joint, and additive frailty.

The frailty model, introduced in the biostatistical literature by Vaupel et al. (1979), and discussed in detail by Hougaard (1984, 1986a, b, 1995), Duchateau and Janssen (2008), and Wienke et al. (2001), accounts for the heterogeneity in baseline. This model is an extension of the proportional hazards model in which the hazard rate function depends upon an unobservable random variable. Subjects may be exposed to different risk levels, even after controlling for known risk factors, because of some relevant unobserved covariates. In a shared frailty model, subjects in the same group share the same frailty value which generates dependence between those subjects who share frailties.

The shared frailty model can be written as follows:

$$\rho_{ik}(t|u_i) = u_i \rho_{ik}(t) = \rho_0(t) \exp(\mathbf{z}'_{ik}\boldsymbol{\beta} + u_i),$$

where $\rho_{ik}(t)$ is the conditional hazard rate function for the kth subject from the ith cluster conditional on u_i, $\rho_0(t)$ is the baseline hazard, $\boldsymbol{\beta}$ is the fixed effects vector of dimension p, $\mathbf{z}_{ik}$ is the vector of covariates, and u_i is the random effect for the ith cluster. Thus, subjects in the same cluster i share the same frailty factor and it is a conditional hazard model, given the u_i. The cluster may represent a family or a single subject for which multiple episodes are observed.

The distribution of u_i may be Gamma, Gaussian, or another distribution. The gamma distribution is often chosen because of its mathematical tractability and because it is widely used. The one-parameter gamma distribution is defined as:

$$f_w(u) = \frac{v^{1/\theta - 1} e^{(u/\theta)}}{\theta^{1/\theta} \Gamma(1/\theta)}$$

with Γ the gamma function and $\mathrm{E}(u) = 1$ and $\mathrm{Var}(u) = \theta$. This means that subjects in class i with $u_i > 1$ are frail (having a higher risk) while subject with $u_i < 1$ are strong (having a lower risk). The parameter θ gives information on the clusters or classes heterogeneity in the population.

5.3 Illustrations

5.3.1 Poisson Regression Data (N = 1000)

Poisson regression is a commonly used count response regression model where events are considered to be of the same kind. Since the model is the '*foundation*' of other recurrent event models, we will look at some of the models' behavior. Few real-life datasets are truly Poisson, where the mean and variance are equal. The vast majority of datasets that are initially thought to be close to Poisson usually have a larger variance than the mean. When this is the case, we say that the Poisson model is overdispersed, which may cause the standard errors of the estimates to be underestimated. When this is the case, a variable may appear to be a significant predictor when in fact it is not. We will illustrate these behaviors by generating data that follows a Poisson regression model, then remove a predictor and see the effect this has on the Poisson model.

We are going to generate $\mathrm{n} = 1000$ standard normally distributed observation for each of 3 independent variables Z_1, Z_2, and Z_3, and then apply the linear equation $\mathbf{Z}'\boldsymbol{\beta} = \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3$ with coefficient $(\beta_1, \beta_2, \beta_3) = (-0.50, -0.50, -0.25)$. After exponentiating $\mathbf{Z}'\boldsymbol{\beta}$ the Poisson variate y is generated using the probability

Table 5.1 General linear Poisson model on y with independent variables Z_1, Z_2, and Z_3

```
Generalized linear models                         No. of obs      =      1,000
Optimization     : ML                             Residual df     =        996
                                                  Scale parameter =          1
Deviance         =  1129.256566                   (1/df) Deviance =   1.133792
Pearson          =  1085.205183                   (1/df) Pearson  =   1.089563
Variance function: V(u) = u                       [Poisson]
Link function    : g(u) = Ln(u)                   [Log]
                                                  AIC             =   2.629619
Log likelihood   = -1310.809415                   BIC             =  -5750.868
------------------------------------------------------------------------------
             |                 OIM
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          z1 |  -.4896427   .0270418   -18.11   0.000    -.5426436   -.4366417
          z2 |  -.4291967   .0286559   -14.98   0.000    -.4853613   -.3730322
          z3 |  -.2327014   .0274627    -8.47   0.000    -.2865273   -.1788755
       _cons |  -.0219569    .034287    -0.64   0.522    -.0891583    .0452444
------------------------------------------------------------------------------
```

Table 5.2 General linear Poisson model on y with independent variables Z_2 and Z_3

```
Generalized linear models                         No. of obs      =      1,000
Optimization     : ML                             Residual df     =        997
                                                  Scale parameter =          1
Deviance         =  1451.594534                   (1/df) Deviance =   1.455962
Pearson          =  1503.530661                   (1/df) Pearson  =   1.508055
Variance function: V(u) = u                       [Poisson]
Link function    : g(u) = Ln(u)                   [Log]
                                                  AIC             =   2.949957
Log likelihood   = -1471.978399                   BIC             =  -5435.437
------------------------------------------------------------------------------
             |                 OIM
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          z2 |  -.4736272   .0284366   -16.66   0.000    -.5293618   -.4178925
          z3 |  -.2235353   .0276194    -8.09   0.000    -.2776683   -.1694023
       _cons |   .1155962   .0311098     3.72   0.000     .0546222    .1765703
------------------------------------------------------------------------------
```

integral transform methods of Kemp and Kemp (1990, 1991) and the method of Kachitvichyanukul (1982) through the Stata software.

The Poisson variate y is next modeled on the three randomly generated independent variables Z_1, Z_2, and Z_3. The results of the analysis are presented in Table 5.1.

Although a sample size of n = 1000 usually is considered to be a decent sample size for a clinical study with one single treatment arm, we will find some interesting outcomes from the analysis results. We find that the estimates of the coefficients (β_1, β_2, β_3) are (−0.4896, −0.4292, −0.2327). All parameter estimates are lower than what we assigned them to be, especially the estimate of β_2 which is −0.4292 instead of −0.50. The Pearson dispersion statistic, defined as the Pearson statistic divided by the model degrees of freedom, would be equal to 1.0 if the model is the '*correct*' one. Here the Pearson statistic is 1.089563, which is about 9% higher than expected.

We will now omit predictor Z_1 and again model the data on the remaining variables. Thus, the Poisson variate y is then modeled on the two randomly generated independent variables Z_2 and Z_3. The results of the analysis are presented in Table 5.2.

We find that the estimates of the coefficients (β_2, β_3) are (−0.4736, −0.2235). Both parameter estimates are still lower than what we assigned them to be, but not worse than from the previous model. The Pearson dispersion statistic has

now increased from 1.089563 to 1.508055. Here the Pearson statistic has notably increased, telling us that the model is overdispersed. The AIC and BIC statistics are also inflated.

5.3.2 *Poisson Regression Data (N = 25,000)*

We will now increase the sample size to 25,000, but everything else will be kept the same. When we use all of the independent variables Z_1, Z_2, and Z_3 in the model (table not shown), we find what we can expect, that the Pearson dispersion statistic is 1.0121, which is very close to 1.0. The estimates of the coefficients (β_1, β_2, β_3) are (-0.4889, -0.5058, -0.2563), i.e., the parameter estimates are very close to what we assigned them to be.

Again, we now omit predictor Z_1 and again model the data on the remaining variables. Thus, the Poisson variate y is then modeled on the two randomly generated independent variables Z_2 and Z_3. The results of the analysis are presented in Table 5.3.

We find that the estimates of the coefficients (β_2, β_3) are (-0.5062, -0.2565). Both parameter estimates are still close to what we assigned them to be, though this is not an indication that the model is appropriate. What has notably changed is that the Pearson dispersion statistic is now about 1.375, telling us that the model is overdispersed. Given the very large dataset of 25,000 observations, the dispersion statistic correctly indicates that the Poisson model is overdispersed. We see that the model obviously does not fit the data.

We will now assume that the variance is proportional rather than equal to the mean, and estimate the scale parameter φ dividing Pearson's chi-squared by its degrees of freedom (df), which gives us the value 1.375. We see that the variance is about 37.5% larger than the mean. This means that we should adjust the standard errors multiplying by 1.173, the square root of 1.375 (see Table 5.4).

Using this procedure of scaling the standard errors we have essentially attributed all the lack of fit to pure error. We can also try to run the Poisson model with the

Table 5.3 General linear Poisson model on y with independent variables Z_2 and Z_3

```
Generalized linear models                         No. of obs      =     25,000
Optimization     : ML                             Residual df     =     24,997
                                                  Scale parameter =          1
Deviance         =  34918.46873                   (1/df) Deviance =   1.396906
Pearson          =  34373.65875                   (1/df) Pearson  =   1.375111
Variance function: V(u) = u                       [Poisson]
Link function    : g(u) = Ln(u)                   [Log]
                                                  AIC             =   2.953739
Log likelihood   =  -36918.7331                   BIC             =  -218216.9
------------------------------------------------------------------------------
             |                 OIM
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          z2 |  -.5062168    .005487   -92.26   0.000    -.5169712   -.4954625
          z3 |  -.2564794   .0054965   -46.66   0.000    -.2672522   -.2457065
       _cons |   .1130332   .0063441    17.82   0.000     .1005989    .1254674
------------------------------------------------------------------------------
```

Table 5.4 General linear Poisson model on y with independent variables Z_2 and Z_3

```
Generalized linear models                         No. of obs      =     25,000
Optimization     : ML                             Residual df     =     24,997
                                                  Scale parameter =          1
Deviance         =  34918.46873                   (1/df) Deviance =   1.396906
Pearson          =  34373.65875                   (1/df) Pearson  =   1.375111
Variance function: V(u) = u                       [Poisson]
Link function    : g(u) = Ln(u)                   [Log]
                                                  AIC             =   2.953739
Log likelihood   =  -36918.7331                   BIC             =  -218216.9
------------------------------------------------------------------------------
             |                 OIM
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          z2 |  -.5062168   .0064344   -78.67   0.000     -.518828   -.4936057
          z3 |  -.2564794   .0064454   -39.79   0.000    -.2691122   -.2438466
       _cons |   .1130332   .0074394    15.19   0.000     .0984522    .1276142
------------------------------------------------------------------------------
(Standard errors scaled using square root of Pearson X2-based dispersion.)
```

robust option to compute standard errors using the robust or '*sandwich*' estimator. Doing so we will get very similar results. In either case, all tests have to be done using Wald's statistic. Likelihood ratio tests are not possible because we are not making full distributional assumptions about the outcome, relying instead on assumptions about the mean and variance.

5.3.3 *Negative Binomial Regression (N = 25,000)*

Since the Poisson model with the two independent variables Z_2 and Z_3 was overdispersed, we will now fit a negative binomial model to the recent data with the same variables Z_2 and Z_3. The results are shown in Table 5.5.

The alpha in Table 5.5 is the variance of the multiplicative random effect. We have overwhelming evidence of overdispersion. For testing hypotheses about the regression coefficients, we can use either Wald tests or likelihood ratio tests.

Table 5.5 Negative Binomial model on y with independent variables Z_2 and Z_3

```
Negative binomial regression                      Number of obs     =     25,000
                                                  LR chi2(2)        =    6437.40
Dispersion     = mean                             Prob > chi2       =     0.0000
Log likelihood = -36210.117                       Pseudo R2         =     0.0816
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          z2 |  -.5068882   .0067312   -75.30   0.000    -.5200811   -.4936952
          z3 |    -.25639   .0066132   -38.77   0.000    -.2693517   -.2434284
       _cons |   .1127922   .0071559    15.76   0.000      .098767    .1268175
-------------+----------------------------------------------------------------
    /lnalpha |  -1.342182   .0367011                     -1.414114   -1.270249
-------------+----------------------------------------------------------------
       alpha |    .261275   .0095891                      .2431408    .2807617
------------------------------------------------------------------------------
LR test of alpha=0: chibar2(01) = 1417.23             Prob >= chibar2 = 0.000
```

Table 5.6 Comparing estimates and standard errors side by side

Variable	poisson Table 3.2.1.	overdisp Table 3.2.2.	nbreg Table 3.3.1.
y			
z2	-.50621684	-.50621684	-.50688815
	.00548703	.00643438	.0067312
z3	-.25647936	-.25647936	-.25639002
	.00549646	.00644543	.00661321
_cons	.1130332	.1130332	.11279225
	.00634412	.00743944	.00715588
lnalpha			
_cons			-1.3421817
			.03670105

5.3.4 Comparing Estimates and Standard Errors

The parameter estimates based on the negative binomial model are not very different from those based on the Poisson regression model. We will now compare the models side by side in Table 5.6.

Both sets of parameter estimates would lead to the same conclusions. Looking at the standard errors reported just below the coefficient estimates, we see that both approaches to overdispersion lead to very similar estimates and that ordinary Poisson regression underestimates the standard errors.

5.3.5 Goodness of Fit

We will evaluate the goodness of fit using the second dataset above with 25,000 observations. One way to compute the deviance of the negative binomial model is to feed the estimate of the variance into the generalized linear model. The deviance statistic is now 1.0741, which tells us that the negative binomial model fits much better than the Poisson model, but still, has a deviance just above the five percent value. One way to model this type of situation is to assume that the data come from a mixture of two populations, one where the counts are always zero, and another population where the count has a Poisson distribution with mean μ. In this, model zero counts can come from either population, while positive counts come only from the second population.

The distribution of the outcome can then be modeled in terms of two parameters, π the probability of '*always zero*', and μ, the mean number of for those not in the '*always zero*' population. A natural way to introduce covariates is to model the logit of the probability π of always zero and the log of the mean μ for those not in the always zero population.

5.3.6 Simulations

Clinical trial simulation studies can be used to assess the impact of many aspects of trial design, conduct, analysis and decision making. Simulation studies can play a vital role in improving the efficiency of drug development within the pharmaceutical industry, but only if they are well designed and conducted. An efficient way of evaluating the properties that different models have for the study design and analysis that we are considering is to use simulations. A number of common software packages make this possible, such as EAST, SAS, Stata, and R.

A comprehensive overview is given of how to use simulations for designing clinical trials and how to analyze the simulated clinical trial data in Ette et al. (2002). A generic template for clinical trials simulations that are typically required by statisticians has been developed by Westfall et al. (2008). Realistic clinical trials datasets are created using a unifying model that allows general correlation structures for endpoint and timepoint data and nonnormal distributions (including time-to-event), and computationally efficient algorithms are presented. The structure allows for patient dropout and noncompliance. A grid-enabled SAS-based system has been developed to implement this model and details are presented summarizing the system development (Westfall et al. 2008, 2010).

For instance, we may use simulations to compare the conditional frailty model and several variance-corrected and frailty models with a known data generating process that exhibits heterogeneity, event dependence, both, and neither. Box-Steffensmeier and De Boef (2006) did this and focused their simulations on the comparison of the three more popular and promising variance-corrected models: the Andersen–Gill, conditional gap time, and conditional elapsed time models, and the basic frailty model estimated with a gamma random effect. They gauged model performance on three dimensions: the bias in the estimated treatment effects as well as in the estimated variance of the random effect, bias in the standard errors, and rate of which the estimated standard errors includes the true parameter. Their simulations suggested that the conditional frailty model can estimate the effects of both sources of correlation simultaneously and retrieve the parameters of the true data generating process better in all four cases. Furthermore, in the simulations they investigated, the conditional frailty model performed similarly to, or better than, the variance-corrected and frailty alternatives. In the case of both heterogeneity and event dependence, only the conditional frailty model performed well. So, in cases where there is a possibility of both, and often we cannot rule either out, the conditional frailty model is recommended.

5.4 Discussion

The foundation of recurrent event analysis is survival analysis has been a common and well-accepted strategy to study treatment effect in a population of patients. During

the last few years, there has been an increasing interest in assessing therapy effect not only by using time to death, but also time to surrogate events such as time to hospitalization. The combined endpoint of time to death and time to disease-related hospitalizations is often analyzed with a time-to-first-event analysis, which has the drawback of waste of information and indistinct handling of two clinically different events.

The analysis of multiple events per subject cannot be approached by a standard Cox model, where the assumption of independence of observations is not valid. In order to account for intra-subject correlation, we have presented the use of marginal and multistate models using a counting process approach for, for instance, the joint analysis of survival and time to disease-related hospitalizations.

In a comparison of common statistical methods for analyzing recurrent event data, the results with each method for lack of bias, efficiency, and robustness for within-subject correlation are not, but depending on the process driving the event counts. In general, the Poisson regression with correction for overdispersion has similar coverage probabilities of confidence intervals, but slightly higher type I error rates compared to the robust Andersen–Gill and negative binomial approaches, which are therefore preferable. Advantages in power for some situations are only at the price of an increased type I error. The negative binomial regression surprisingly produces results similar to those of the Andersen–Gill approach, even when the distribution is not homogeneously Poisson. On the other hand, for homogeneous Poisson processes, the Andersen–Gill approach does not lose efficiency in comparison with the perfectly fitting negative binomial regression model (Jahn-Eimermacher et al. 2015). The demonstrated comparability of the Andersen–Gill approach and negative binomial regression for Poisson processes supports the findings of Metcalfe and Thompson (2006). The results are in agreement with the data example presented by Guo et al. (2008), in which trial results from an Andersen–Gill model were similar to those from Poisson regression.

For the conditional model not derived from the Poisson process, with all the investigated methods, estimation of a zero treatment effect and its standard error may be considered as acceptable, and thus be applicable to hypothesis testing. However, the effect estimates are biased, whatever method is used. All of the investigated methods are not applicable if the independent increment assumption is violated. For a specific application, this assumption, therefore, must be checked by appropriate sensitivity analyses. So, results could be compared with those of the conditional model of Prentice et al. (1981) or the marginal model of Wei et al. (1989). However, these approaches also have sources of bias as demonstrated by Therneau and Grambsch (2000) and Kelly and Lim (2000) and, furthermore, the applicability of the marginal model to recurrent failure time data is discussed critically in Metcalfe and Thompson (2007).

We found no advantages in performance with Poisson regression as compared with the Andersen–Gill approach, which allows more complex analyses and may, therefore, be preferable. The Poisson regression remains applicable when only aggregated event counts are available or when the actual time of occurrence of an event cannot be determined. Dean and Balshaw (1997) demonstrated for nonhomogeneous Pois-

son processes that treatment effects can be efficiently estimated based on aggregated count data as long as censoring is balanced between treatment groups.

Standard errors might be substantially underestimated with all the methods examined if within-subject correlation is not accounted for, in accordance with previous findings (Glynn and Buring 1996, Therneau and Hamilton 1997, Metcalfe and Thompson 2006). Robust variance estimation can be used to adjust for the simulated degree of within-subject-correlation, however, in rare cases, data may be even more highly correlated (Thall 1988). In those situations, the robust methods may also fail to prevent type I error from increasing to unacceptable levels.

The use of a gamma distribution for the random effect is common in the literature (Stukel 1993; Metcalfe and Thompson 2006; Thomsen and Parner 2006). Regression parameter estimation in a gamma frailty model seems to be robust to frailty distribution misspecification as Hsu et al. (2007) demonstrated for single event data in cohort and case-control family trials. Kelly and Lim (2000), Therneau and Grambsch (2000) and Metcalfe and Thompson (2006) used realizations from normal and uniform distributions, with which the Andersen–Gill method underestimated treatment effects.

Finally, the most appropriate model should be chosen based on the anticipated nature and structure of the data.

References

Aalen, O. O., Borgan, Ø. & Gjessing, H. K. (2008). *Survival and event history analysis: A process point of view*. Springer.

Altman, D. G., & Royston, P. (2000). What do we mean by validating a prognostic model? *Statistics in Medicine, 19,* 453–473.

Andersen, P. K., Abildstrom, S. Z., & Rosthøj, S. (2002). Competing risks as a multi-state model. *Statistical Methods in Medical Research, 11,* 203–215.

Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. (1993). *Statistical models based on counting processes*. New York: Springer-Verlag.

Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. (1997). *Statistical models based on counting processes* (Corrected ed.). New York: Springer.

Andersen, P. K., & Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics, 10,* 1100–1120.

Andersen P. K., & Keiding, N. (2002). Multi-state models for event history analysis. *Statistical Methods in Medical Research*, *11*, 91.

Blossfeld, H., & Rohwer, G. (1995). *Techniques of event history modeling*. New Jersey: Lawrence Erlbaum.

Box-Steffensmeier, J. M., & De Boef, S. (2006). Repeated events survival models: The conditional Frailty model. *Statistics in Medicine, 25,* 3518–3533.

Brenner, H., & Gefeller, O. (1997). Deriving more up-to-date estimates of long-term patient survival. *Journal of Clinical Epidemiology, 50,* 211–216.

Breslow, N. E. (1975). Analysis of survival data under the proportional hazards model. *International Statistical Review/Revue Internationale de Statistique*. *43*(1), 45–57.

Coleman, M. P., Forman, D., Bryant, H., Butler, J., Rachet, B., Maringe, C., Nur, U., et al. (1995–2007). Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK,

(the International Cancer Benchmarking Partnership): An analysis of population-based cancer registry data. *Lancet, 377*: 127–138.
Commenges, D. (1999). Multi-state models in epidemiology. *Lifetime Data Analysis, 5,* 315–327.
Cook, R. J., & Lawless, J. F. (2002). Analysis of repeated events. *Statistical Methods in Medical Research, 11,* 141–166.
Cook, R. J., & Lawless, J. F. (2007). *The statistical analysis of recurrent events*. Springer.
Courgeau, D., & Lelièvre, E. (1992). *Event history analysis in demography*. Oxford: Clarendon.
Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B, 34*(2), 187–220.
Crowther, M. J. (2016). *MULTISTATE: Stata module to perform multi-state survival analysis*. https://ideas.repec.org/c/boc/bocode/s458207.html.
Dean, C. B., & Balshaw, R. (1997). Efficiency lost by analyzing counts rather than event times in Poisson and overdispersed Poisson regression models. *Journal of the American Statistical Association, 92,* 1387–1398.
Duchateau, L., & Janssen, P. (2008). *The Frailty model*. Springer.
Efron, B. (1974). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association., 72*(359), 557–565.
Ette, E. I., Godfrey, C. J., Ogenstad, S., & Williams, P. (2002). Analysis of simulated clinical trials. In *Simulation for designing clinical trials: A pharmacokinetic-pharmacodynamic modeling perspective.* Marcel-Dekker, IBSN: 0-8247-0862-8.
Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association, 94,* 496–509.
Fuchs, H. J., Borowitz, D. S., Christiansen, D. H., Morris, E. M., Nash, M. L., Ramsey, B. W., Rosenstein, B. J. et al. (1994). Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. The Pulmozyme Study Group. New England Journal of Medicine, *331*(10):637–42.
Gamel, J. W., & Vogel, R. L. (2001). Non-parametric comparison of relative versus cause-specific survival in Surveillance, Epidemiology and End Results (SEER) programme breast cancer patients. *Statistical Methods in Medical Research, 10,* 339–352.
Ghosh, D., & Lin, D. Y. (2000). Nonparametric analysis of recurrent events and death. *Biometrics, 56*(2), 554–562.
Glynn, R. J., & Buring, J. E. (1996). Ways of measuring rates of recurrent events. *British Medical Journal, 312,* 364–367.
Guo, Z., Gill, T. M., & Allore, H. G. (2008). Modeling repeated time-to-event health conditions with discontinuous risk intervals. *Methods of Information in Medicine, 47,* 107–116.
Gutierrez, R. G. (2002). Parametric frailty and shared frailty survival models. *Stata Journal, 2,* 22–44.
Hilbe, J. M. (2007). *Negative binomial regression*. Cambridge University Press.
Hougaard, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika, 71,* 75–83.
Hougaard, P. (1986a). A class of multivariate failure time distributions. *Biometrika, 73,* 671–678.
Hougaard, P. (1986b). Survival models for heterogeneous populations derived from stable distributions. *Biometrika, 73*, 397–96.
Hougaard, P. (1995). Frailty models for survival data. *Lifetime Data Analysis, 1,* 255–273.
Hougaard, P. (1999). Multi-state models: A review. *Lifetime Data Analysis, 5,* 239–264.
Hougaard, P. (2000). *Analysis of multivariate survival data*. New York: Springer.
Hsu, L., Gorfine, M., & Malone, K. (2007). On robustness of marginal regression coefficient estimates and hazard functions in multivariate survival analysis of family data when the frailty distribution is mis-specified. *Statistics in Medicine, 26,* 4657–4678.
Jahn-Eimermacher, A., Ingel, K., Ozga, A. K., Preussler, S., & Binder, H. (2015). Simulating recurrent event data with hazard functions defined on a total time scale. *BMC Medical Research Methodology, 15,* 16.

Kachitvichyanukul, V. (1982). *Computer generation of poisson, binomial, and hypergeometric random variables*. Ph.D. thesis, Purdue University.

Kelly, P. J., & Lim, L. L. (2000). Survival analysis for recurrent event data: An application to child-hood infectious diseases. *Statistics in Medicine, 19,* 13–33.

Kemp, A. W., & Kemp, C. D. (1990). A composition-search algorithm for low-parameter Poisson generation. *Journal of Statistical Computation and Simulation, 35,* 239–244.

Kemp, C. D., & Kemp, A. W. (1991). Poisson random variate generation. *Applied Statistics, 40,* 143–158.

Lambert, P. C., & Royston, P. (2009). Further development of flexible parametric models for survival analysis. *Stata Journal, 9,* 265–290.

Li, Q., & Lagakos, S. (1997). Use of the Wei-Lin-Weissfeld method for the analysis of a recurring and terminating event. *Statistics in Medicine, 16,* 925–940.

Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: The marginal approach. *Statistics in Medicine, 13,* 2233–2247.

Lin, D. Y., & Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association, 84,* 1074–1078.

Lunn, M., & McNeil, D. (1995). Applying Cox regression to competing risks. *Biometrics, 51,* 524–532.

Marso, S. P., Daniels, G. H., Brown-Frandsen, K., Kristensen, P., Kristensen, P., Mann, J. F. E., et al. (2016). Liraglutide and cardiovascular outcomes in type 2 diabetes. *New England Journal of Medicine, 375,* 311–322.

Metcalfe, C., & Thompson, S. G. (2006). The importance of varying the event generation process in simulation studies of statistical methods for recurrent events. *Statistics in Medicine, 25,* 165–179.

Metcalfe, C., & Thompson, S. G. (2007). Wei, Lin and Weissfeld's marginal analysis of multivariate failure time data: Should it be applied to recurrent events outcome? *Statistical Methods in Medical Research, 16,* 103–122.

Møller, H., Sandin, F., Bray, F., Klint, A., Linklater, K. M., Purushotham, A., et al. (2010). Breast cancer survival in England, Norway and Sweden: A population-based comparison. *International Journal of Cancer, 127,* 2630–2638.

Prentice, R. L., Williams, B. J., & Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika, 68,* 373–379.

Royston, P., & Lambert, P.C. (2011). *Flexible parametric survival analysis using Stata: Beyond the Cox model*. Stata-Press.

Royston, P., & Parmar, M. K. B. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine, 21,* 2175–2197.

Sagara, I., Giorgi, R., Doumbo, O. K., Piarroux, R., & Gaudart, J. (2014). Modelling recurrent events: Comparison of statistical models with continuous and discontinuous risk intervals on recurrent malaria episodes data. *Malaria Journal, 13,* 293.

Stukel, T. A. (1993). Comparison of methods for the analysis of longitudinal interval count data. *Statistics in Medicine, 12,* 1339–1351.

Thall, P. F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics, 44,* 197–209.

Therneau, T. M. (1997). Extending the Cox model. In *Proceedings of the First Seattle Symposium in Biostatistics*. New York: Springer.

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling survival data*. New York: Springer.

Therneau, T. M., & Hamilton, S. A. (1997). rhDNase as an example of recurrent event analysis. *Statistics in Medicine Banner, 16,* 2029–2047.

Thomsen, J. L., & Parner, E.T. (2006). Methods for analysing recurrent events in health care data. Examples from admissions in Ebeltoft Health Promotion project. *Family Practice, 23*, 407–413.

Tsiatis, A. A. (2005). Competing risks. In P. Armitage & T. Colton (Eds.), *Encyclopedia of Biostatistics* (2nd ed., pp. 1025–1035). Hoboken, NJ: Wiley.

Vaupel, J. W., Manton, K., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography, 16,* 439–454.

Wei, L. J., Lin, D. Y., & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association., 84,* 1065–1073.

Westfall, P. H., Dmitrienko, A., DeSouza, C., & Tomoiaga, A. (2010). Clinical trials simulation: A publicly available, grid-enabled, GUI-Driven SAS® system. SAS Global Forum 2010, Paper 180–2010.

Westfall, P. H., Tsai, K., Ogenstad, S., Tomoiaga, A., Moseley, S., & Lu, Y. (2008). Clinical trials simulation: A statistical approach. *Journal of Biopharmaceutical Statistics, 18*(4), 611–630.

Wienke, A., Holm, H., Skytthe, A., & Yashin, A. (2001). The heritability of mortality due to heart diseases: A correlated Frailty model applied to Danish twins. *Twin Research, 4*(4), 266–274.